

# Machine Learning Week 4. Neural Networks.

## 1. Non-linear hypothesis.

- non-linear classification.  $\rightarrow$  a lot of terms.

· not scalable,  $O(k^2)$ ,  $O(k^3)$

- application: computer vision

- labelled training set  $\Rightarrow$  train model.

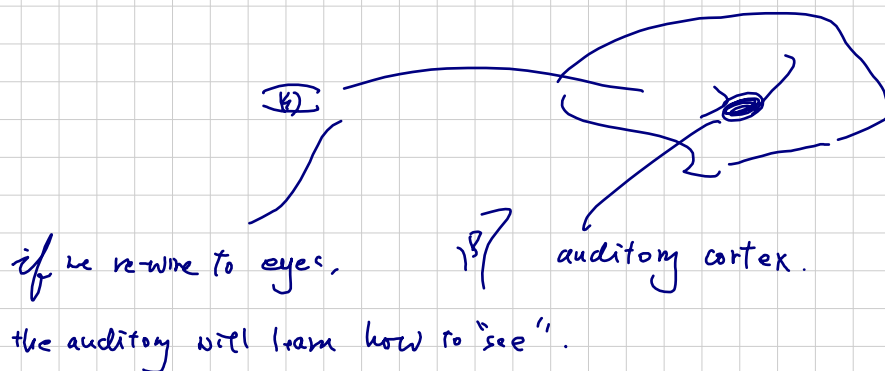
- neural network is useful for non-linear, large scale classification

## 2. Neural networks

· origin  $\rightarrow$  "neuromorphic" to mimic how the brain works.

· recent resurgence: computer (hardware) advancement.

· The "one learning algorithm" hypothesis

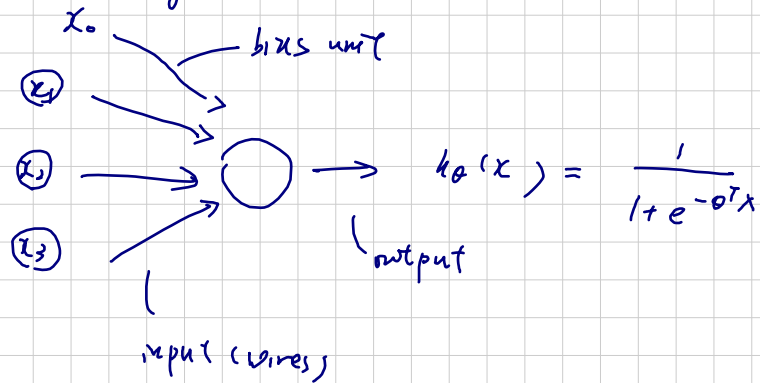


"neural rewiring"  $\rightarrow$  there exists "one algorithm" that can do generalized learning.

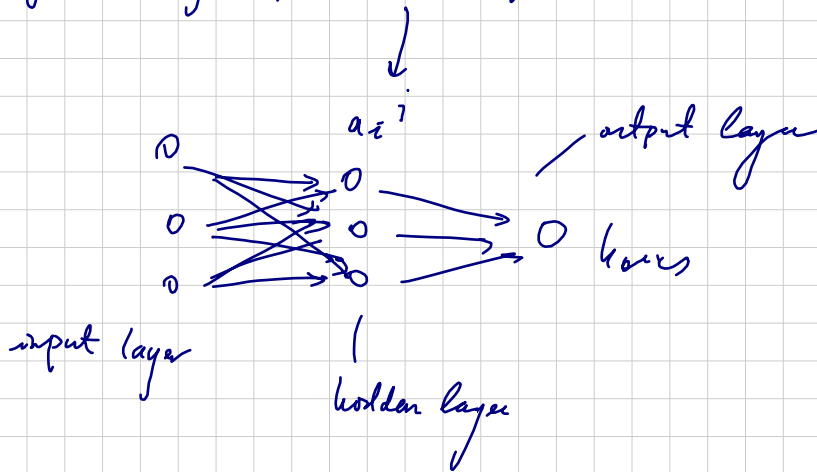
### 3. Neural network model representation

- interneuron communication - sends electric pulse

- Neuron model: logistic unit



sigmoid. (logistic) activation function



#### 4. Model Representation II.

Forward Propagation  $\rightarrow$  vectorized implementation.

$$a_1^{(2)} = g \left( \Theta_{1,0}^{(1)} x_0 + \Theta_{1,1}^{(1)} x_1 + \Theta_{1,2}^{(1)} x_2 + \Theta_{1,3}^{(1)} x_3 \right)$$

$$\Rightarrow a_1^{(2)} = z_2^{(2)}$$

$$x = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad z^{(2)} = \begin{pmatrix} z_1^{(2)} \\ z_2^{(2)} \\ z_3^{(2)} \end{pmatrix} = \Theta^{(1)} (x) \quad \text{define}$$

$$= \Theta^{(1)} \cdot a^{(1)}$$

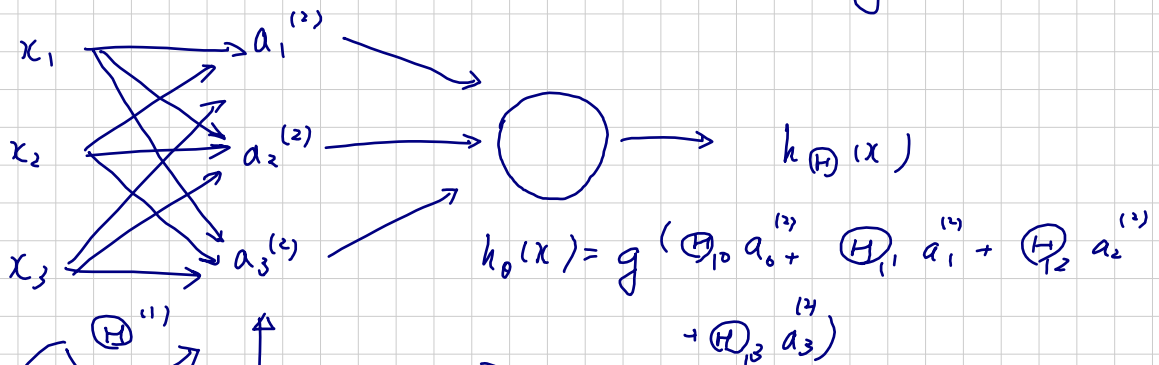
$$a^{(2)} = g(z^{(2)})$$

$$\text{Add. } a_0^{(2)} = 1, \quad a^{(2)} \in \mathbb{R}^{(4)}$$

$\mathbb{R}^3 \rightarrow \mathbb{R}^3$   
element wise.

$$z^{(3)} = \Theta^{(2)} a^{(2)}, \quad h_{\Theta}(x) = a^{(3)} = g(z^{(3)})$$

Neural network learning its own features : flexibility to learn features first [hidden layer] before feeding into logistic regression



Mapping function  
(original  $\rightarrow$  processed)  
features.

"processed" features that feed into the original logistic regression.

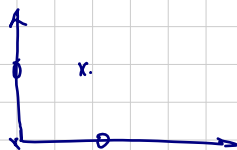
standard logistic model.

other Network Architecture.

- multiple hidden layers  $\rightarrow$  complexity, non-linearity.  $\uparrow$

## 6. Examples and Intuition (1)

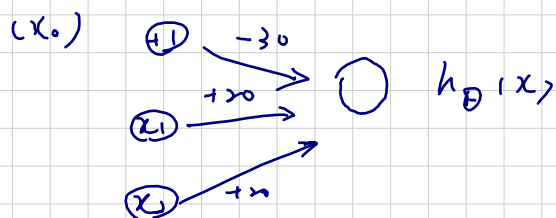
Non-linear classification example. XOR / XNOR.



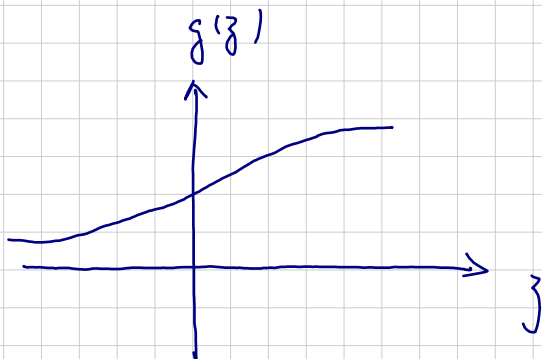
$$y = x_1 \text{ XOR } x_2$$

simple example.

$$x_1, x_2 \in \{0, 1\}$$



$$h_{\Theta}(x) = g \left( \underbrace{-30}_{\Theta_{10}^{(1)}} + \underbrace{20}_{\Theta_{11}^{(1)}} x_1 + \underbrace{20}_{\Theta_{12}^{(1)}} x_2 \right)$$



$x_1$	$x_2$
0	0
0	1
1	0
1	1

$x_1 \text{ \& } x_2$

$$h_{\Theta}(x) \downarrow$$

$$g(-30) = 0$$

$$g(-10) = 0$$

$$g(-10) = 0$$

$$g(10) = 1$$

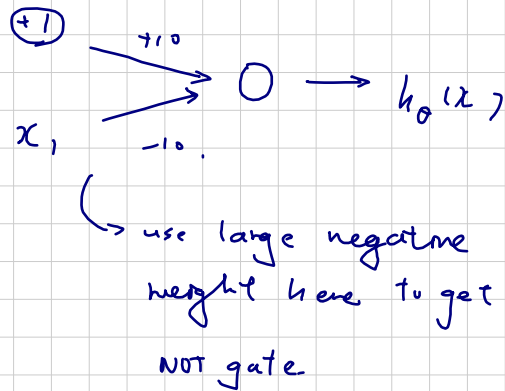
consider.  $\Theta^{(1)} = \begin{pmatrix} -10 \\ 20 \\ 20 \end{pmatrix}$

$x_1$	$x_2$	$z$	$g(z)$
0	0	-10	0
0	1	10	1
1	0	10	1
1	1	30	1

$x_1$  or  $x_2$

## 6. Example and Intuition. (7)

- Not gate.

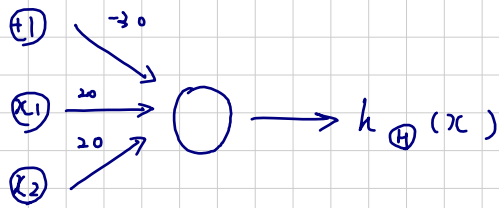


$$h_{\Theta}(x) = g(10 - 20x_1)$$

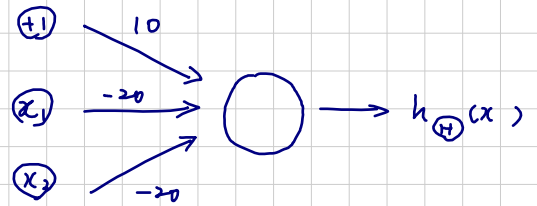
$x_1$	$z$	$g$
0	10	1
1	-10	0

- ex.  $(\text{NOT } x_1) \text{ AND } (\text{NOT } x_2) = 1 \Leftrightarrow x_1 = 0, x_2 = 0.$

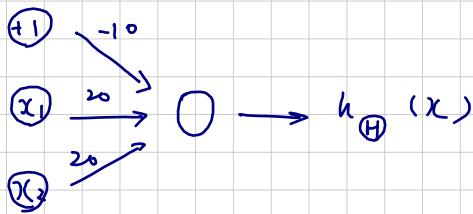
## Summary



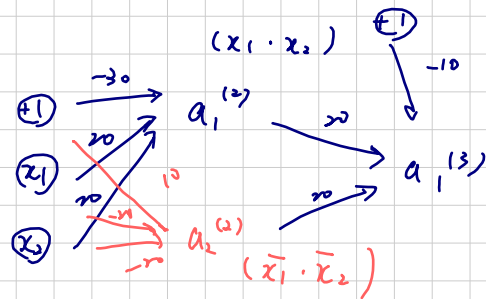
$x_1 \text{ AND } x_2$



$\bar{x}_1 \text{ AND } \bar{x}_2$



$x_1 \text{ OR } x_2$



$x_1 \text{ XOR } x_2$

$$\overline{x_1 \oplus x_2} = x_1 x_2 + \bar{x}_1 \bar{x}_2$$

## 7. Multiclass classification

- multiple output units. (each unit classifies true/false on each bit)